Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal. Vol. 2, Issue 1 (2009), 67 – 76
ISSN 2070-5948, DOI 10.1285/i20705948v2n1p67

A SIMPLE METHOD FOR THE DETECTION OF OUTLIERS

Taichiro Ueda*

Mitsubishi Electric Computer Systems Corporation, Tokyo.

Received 17 August 2009; Accepted 1 September 2009
Available online 15 September 2009

Abstract**: *This article presents a simple and efficient method to detect multiple outliers using a modification of the Akaike's Information Criterion. The proposed method is exemplified using various examples and implemented in the statistical package R. The core features of the method are that it depends on simple computations, it does not depend on traditional p values, and it signals the absence of outliers. The proposed procedure is applicable to analyse data in all sciences.*

Keywords: Masking effect, outliers, AIC, Stirling's formula.

1. Introduction

Detecting outliers (e.g., data from a different population, sampling error, etc.) is an important issue in data analysis and effective methods to deal with them are always in demand. There have already been proposed various statistical methods for the detection of outliers (e.g., Barnett & Lewis, 1994; Grubbs, 1969; Johnson, McGuire, & Milliken, 1978) but these methods present some problems. For example, the computations used when there is only one outlier are different

* Reference in original language: 上田, 太一郎. (1996). 複数外れ値の簡易検出法. *応用統計学, 25* (1), 17-26. [Ueda, T. (1996). Simple method for the detection of outliers. *Japanese Journal of Applied Statistics, 25* (1), 17-26.] Citation in references for the present translation:
Ueda, T. (2009). A simple method for the detection of outliers. *Electronic Journal of Applied Statistical Analysis, 1 (*2009), 67-76. (F. Marmolejo-Ramos & S. Kinoshita, Trans.) (Original work published in 1996)
In the text, cite the original publication date and the date of translation, i.e., Ueda, 1996/2009. Further questions regarding this article can be addressed to fernando.marmolejoramos@adelaide.edu.au. This translation is dedicated to the memory of the author of this work, Dr. Taichiro Ueda. F.M-R thanks Kyoko Katayama, Yuka Toyama, Daisuke Fujiki, and Kimihiro Noguchi for reviewing the accuracy and readability of this translation.

** The original version of this paper does not have an abstract, only three key words. Therefore, an abstract was written for the present translation and an extra key word was added (i.e., outliers). Including an abstract and more key words facilitates indexation.

from those used when there are two outliers. Also, conclusions about the influence of outliers will be different depending on the level of significance adopted (e.g., 5%, 1%, etc).
Similarly, masking effects, i.e., the effect of multiple outliers sustaining other outliers, need to be tackled too. To deal with multiple outliers Kitagawa (1979) proposed the application of the AIC method (Akaike's Information Criterion) as mentioned in Beckman and Cook (1983). The appeal of the AIC method is that it is not affected by the number of outliers and treats all outliers with a unified method; however, it requires complex computations. A simpler method for the detection of outliers is proposed in this paper and it is accompanied by some examples.

2. Definition of statistics and simple detection method of outliers

Kitagawa's (1979) proposal consists of an application of the AIC to the detection of outliers. The simple detection statistic proposed in this paper is based on Kitagawa's (1979) application, where AIC approximates:

-2 (Normal observations' value log likelihood) + 2 (number of parameters)

Outliers are modelled using *f* as a density function of the normal distribution:

$$n!\prod_{J=1}^{n} f(x_j)$$

Thus, AIC becomes:

$$AIC = 2n\log\hat{\sigma} - 2\log n! + 2s \qquad (1)$$

Where $n+s$ is the number of observations, $\hat{\sigma}$ is the standard deviation[1], and s is the number of outlier candidates (s= 0, 1, 2, 3,…). Table 1 shows the values for log $n!$ and log $n!/n$ with respect to n.
Following Table 1, a rough approximation suggests that when $n \approx 5\sim9$, $log\, n! \approx n$. Then using the right side of equation (1), $2s - 2\, log\, n! \approx 2s - 2n = 2s - 2$ (the number of observations - s) = $4s$ - constant. Similar calculations yield approximations for $log\, n!/n$ as given in Table 2 when n ranges from 10 to 2000.

[1] In the original article it is not said what sigma stands for. However by looking at the formula used in the Appendix is it becomes obvious that $\hat{\sigma}$ is the standard deviation (see also footnote 5).

Table 1. Values for log *n*! and log *n*!/*n* with respect to *n*

n	$\log n!$	$\log n!/n$	n	$\log n!$	$\log n!/n$
3	1.8	0.6	120	457.8	3.8
4	3.2	0.8	130	506.1	3.9
5	4.8	1.0	140	555.2	4.0
6	6.6	1.1	150	605.0	4.0
7	8.5	1.2	160	655.5	4.1
8	10.6	1.3	170	706.6	4.2
9	12.8	1.4	180	758.2	4.2
10	15.1	1.5	190	810.5	4.3
11	17.5	1.6	200	863.2	4.3
12	20.0	1.7	250	1134.0	4.5
13	22.5	1.7	300	1414.9	4.7
14	25.2	1.8	350	1704.1	4.9
15	27.9	1.9	400	2000.5	5.0
16	30.7	1.9	450	2303.1	5.1
17	33.5	2.0	500	2611.3	5.2
18	36.4	2.0	550	2924.5	5.3
19	39.3	2.1	600	3242.3	5.4
20	42.3	2.1	650	3564.2	5.5
21	45.4	2.2	700	3890.0	5.6
22	48.5	2.2	750	4219.3	5.6
23	51.6	2.2	800	4552.0	5.7
24	54.8	2.3	850	4887.7	5.8
25	58.0	2.3	900	5226.5	5.8
26	61.3	2.4	950	5568.0	5.9
27	64.6	2.4	1000	5912.1	5.9
28	67.9	2.4	1050	6258.8	6.0
29	71.3	2.5	1100	6607.8	6.0
30	74.7	2.5	1150	6959.1	6.1
31	78.1	2.5	1200	7312.6	6.1
32	81.6	2.5	1250	7668.1	6.1
33	85.1	2.6	1300	8025.7	6.2
34	88.6	2.6	1350	8385.1	6.2
35	92.1	2.6	1400	8746.5	6.2
36	95.7	2.7	1450	9109.6	6.3
37	99.3	2.7	1500	9474.4	6.3
38	103.0	2.7	1550	9840.9	6.3
39	106.6	2.7	1600	10209.0	6.4
40	110.3	2.8	1650	10578.7	6.4
50	148.5	3.0	1700	10949.9	6.4
60	188.6	3.1	1750	11322.6	6.5
70	230.4	3.3	1800	11696.6	6.5
80	273.7	3.4	1850	12072.1	6.5
90	318.2	3.5	1900	12449.0	6.6
100	363.7	3.6	1950	12827.1	6.6
110	410.3	3.7	2000	13206.5	6.6

Table 2. Approximate formulas with respect to size of *n*

n	$\log n!/n$	$2s - 2\log n!$	AIC
5 to 9	1	4*s* - constant	$2n\log\sigma + 4s$
10 to 28	2	6*s* - constant	$2n\log\sigma + 6s$
29 to 82	3	8*s* - constant	$2n\log\sigma + 8s$
83 to 230	4	10*s* - constant	$2n\log\sigma + 10s$
231 to 600	5	12*s* - constant	$2n\log\sigma + 12s$
601 to 1715	6	14*s* - constant	$2n\log\sigma + 14s$
1716 to 2000	7	16*s* - constant	$2n\log\sigma + 16s$

The values $4s, 6s, 8s$, ... in Table 2 are correction terms which depend on n. For example, when the number observations in a sample ($n + s$) is 10, the correction term of AIC could be either $4s$ or $6s$ which is not good. To correct for this, it is proposed an approximation to the AIC using the formula, $\sqrt{2} \cdot s \cdot \frac{\log n!}{n}$, which depends on n and shows a smooth behaviour. Table 3 shows how this approximation performs regarding AIC; the results are similar. Accordingly, the outlier detection statistic U_t is defined as follows.

$$U_t = \frac{1}{2} AIC = n\log\hat{\sigma} - \sqrt{2}\cdot s\cdot\frac{\log n!}{n} \qquad (2)$$

Table 3. Approximate values for the correction term

	s							
	Half value of correction term $\frac{1}{2}AIC$				*Formula for correction term* $\sqrt{2}\cdot s\cdot\frac{\log n!}{n}$			
n	*s* =1	*s* =2	*s* =3	*s*=4	*s* =1	*s* =2	*s* =3	*s*=4
10	2	4	6	8	2.1	4.3	6.4	8.5
20	3	6	9	12	3.0	6.0	9.0	12.0
50	4	8	12	16	4.2	8.4	12.6	16.8
100	5	10	15	20	5.1	10.3	15.4	20.6
150	5	10	15	20	5.7	11.4	17.1	22.8
200	5	10	15	20	6.1	12.2	18.3	24.4
300	6	12	18	24	6.7	13.3	20.0	26.7
700	7	14	21	28	7.9	15.7	23.6	31.4
1000	8	16	24	32	8.4	16.7	25.1	33.4
2000	8	16	24	32	9.3	18.7	28.0	37.4

In order to find outliers, all that is required is to determine a combination of data points that gives the smallest statistic U_t. An example is useful to clarify this. Let's assume there is a data set composed of the following observations, 5.71, 6.57, 7.29, 8.06, and 13.32 (taken from Takeuchi, 1980). First, subtract the mean from each observation and divide it by the data set's standard deviation (i.e., compute *z* scores). The obtained values are - 0.828, - 0.541, - 0.300, - 0.043, and 1.712 respectively. If no outlier is expected, then $n = 5$, and $s = 0$. In such case,

$U_t = 5\log\hat{\sigma} + 0 = -0.56$. However, if it is assumed that the last observation (13.32, which correspond to the *z* value of 1.712) is an outlier, then $n = 4$, and $s = 1$. In such case, $U_t = 4\log\hat{\sigma} + 2 \times 1 = -3.83$. Computations of U_t for all possible combinations of potential outliers using Takeuchi's data are shown in Table 4. According to Table 4, the only outlier is13.32 and its identification threw the lowest U_t. Note that Takeuchi (1980) also used 13.32 as the only outlier in his data set.

Table 4. Values of U_t when some observations are removed

	Largest value removed		
Smallest value removed	*None*	*13.32*	*13.32 8.06*
None	-0.56	<u>-3.83</u>	-2.94
5.71	0.63	-3.12	-2.86
5.71 6.57	1.33	-2.72	--

Note. The minimum U_t obtained is underlined (Same approach is used in the following tables)

3. Application example

In the following examples, U_t is based on the value of the data divided by the standard deviation.

3.1 Data 1 from Grubbs (1969)

It can be seen from the data in the Table 5 that 2.02 and 2.22 could be potential outliers. According to Table 5, the minimum U_t obtained is -1.77, which occurs when 2.02 and 2.22 are signalled as outliers. Kitagawa (1979) also determined observations 2.02 and 2.22 as outliers.

Table 5. Data 1 from Grubbs (1969) and U_t obtained.

Data (sample size = 10): 2.02, 2.22, 3.04, 3.23, 3.59, 3.73, 3.94, 4.05, 4.11, 4.13

	Largest value removed			
Smallest value removed	*None*	*4.13*	*4.13 4.11*	*4.13 4.11 4.05*
None	-0.53	1.49	3.13	4.35
2.02	-0.28	1.70	3.25	4.33
2.02, 2.22	<u>-1.77</u>	0.21	1.65	2.48
2.02, 2.22, 3.04	-1.30	0.57	1.82	2.33

3.2 Data 2 from Grubbs (1969)

Data is shown in Table 6. According to Table 6, the minimum U_t obtained is -2.50, which occurs when -1.40 and 1.01 are signalled as outliers. Kitagawa (1979) signalled -1.40, 1.01, and 0.63 as the first list of candidate outliers, his second candidate list was - 1.40 and 1.01.

Table 6. Data 2 from Grubbs (1969) and U_t obtained.

Data (sample size = 15): -1.40, -0.44, -0.30, -0.24, -0.22, -0.15, -0.13, 0.06, 0.10, 0.18, 0.20, 0.39, 0.48, 0.63, 1.01

	Largest value removed			
Smallest value removed	*None*	*1.01*	*1.01* *0.63*	*1.01* *0.63* *0.48*
None	-0.52	0.53	2.38	4.30
-1.40	-2.30	-2.50	-1.38	-0.12
-1.40, -0.44	-0.20	-0.62	0.31	1.35
-1.40, -0.44, -0.30	2.04	1.49	2.28	3.15

3.3 Data affected by masking

Takeuchi and Ohashi (1981) reported a data set which exemplifies the effect of masking (see Table 7). According to the method proposed in this paper, observations -67 and -48 are the only outliers. Takeuchi and Ohashi (1981) also identified observations -67 and -48 as outliers. A closer look at Table 7 enables to identify the outliers even in cases in which data presents masking effects.

Table 7. Data affected by masking and U_t obtained.

Data (sample size = 15): -67, -48, 6, 8, 14, 16, 23, 24, 28, 29, 41, 49, 56, 60, 75

	Largest value removed			
Smallest value removed	*None*	*75*	*75* *60*	*75* *60* *56*
None	-0.52	1.35	3.46	5.37
-67	-1.21	0.36	2.33	4.06
-67, -48	-2.91	-2.19	-0.79	0.04
-67, -48, 6	-0.58	0.003	1.26	1.92

3.4 Cases in which there is no outlier

The following example represents the case in which no outliers seem to be present (see Table 8). According to Table 8, $U_t = -0.53$ which is the minimum value obtained when no observation is signalled as outlier.

Table 8. Situation in which no outlier is present and U_t obtained.

Data (sample size = 10): 5.4, 5.4, 5.5, 5.7, 5.8, 5.9, 6.0, 6.1, 6.3, 6.4

	Largest value removed		
Smallest value removed	*None*	*6.4*	*6.4* *6.3*
None	-0.53	0.45	1.02
-67	1.03	1.88	2.31
-67, -48	1.88	2.53	2.63

3.5 Data in which the number of samples is 65 and masking effects exist

Table 9 shows a data set used by Washio and Ohashi (1989) to describe the length of skis. In this data set, minimum U_t is -2.18, when observations 160 and 161 are signalled as outliers.

Table 9. Skis data and U_t o obtained.

Data (sample size = 65): 160, 161, 170, 170, 170, 170, 175, 176, 180, 180, 180, 181, 182, 183, 183, 185, 185, 188, 188, 188, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 190, 191, 191, 191, 191, 191, 191, 191, 191, 191, 194, 195, 195, 197, 200, 200, 201, 202, 202, 203

	Largest value removed				
Smallest value removed					*203*
				203	*202*
			203	*202*	*202*
	None	*203*	*202*	*202*	*201*
None	-0.50	2.96	6.50	9.91	13.41
160	-0.98	2.35	5.76	9.03	12.40
160, 161	-2.18	0.96	4.19	7.25	10.41
160, 161, 170	-0.43	2.59	5.73	8.66	11.70
160, 161, 170, 170	0.83	3.72	6.71	9.48	12.36

3.6 Data set used in Rosner (1983)

Table 10 shows the data set used in Rosner (1983). Both the present and Rosner's method identify observations 6.01, 5.42, and 5.34 as outliers.

Table 10. Rosner's data set and U_t obtained.

Data (sample size = 54): -0.25, 0.68, 0.94, 1.15, 1.20, 1.26, 1.26, 1.34, 1.38, 1.43, 1.49, 1.49, 1.55, 1.56, 1.58, 1.65, 1.69, 1.70, 1.76, 1.77, 1.81, 1.91, 1.94, 1.96, 1.99, 2.06, 2.09, 2.10, 2.14, 2.15, 2.23, 2.24, 2.26, 2.35, 2.37, 2.40, 2.47, 2.54, 2.62, 2.64, 2.90, 2.92, 2.92, 2.93, 3.21, 3.26, 3.30, 3.59, 3.68, 4.30, 4.64, 5.34, 5.42, 6.01

	Largest value removed				
Smallest value removed					*6.01*
				6.01	*5.42*
			6.01	*5.42*	*5.34*
	None	*6.01*	*5.42*	*5.34*	*4.64*
None	-0.50	-1.21	-1.22	-2.12	-1.61
-0.25	1.76	0.71	0.34	-1.12	-1.10
-0.25, 0.68	5.38	4.23	3.75	2.10	1.94
-0.25, 0.68, 0.94	9.23	8.03	7.48	5.72	5.45
-0.25, 0.68, 0.94, 1.15	13.23	12.01	11.43	9.61	9.27

3.7 Situation when outliers are added to a normal distribution

A normally distributed data set composed of 100 random observations with a mean of 0 and a standard deviation of 1 was generated. Then 5 outliers were added. The added outliers are -4.00, 3.60, 4.10, 5.20, and 5.70. Table 11 presents the contaminated data set.

Table 11. Generated data and U_t obtained.

Data (sample size = 105): -4.00, -2.83, -2.79, -2.42, -2.40, -2.83, -2.12, -1.93, -1.60, -1.51, -1.27, -1.26, -1.20, -1.19, -1.11, -1.08, -1.06, -1.05, -1.03, -1.01, -0.99, -0.98, -0.93, -0.93, -0.91, -0.90, -0.86, -0.82, -0.78, -0.77, -0.75, -0.71, -0.71, -0.60, -0.59, -0.58, -0.57, -0.43, -0.41, -0.38, -0.36, -0.36, -0.35, -0.33, -0.32, -0.31, -0.29, -0.29, -0.27, -0.24, -0.24, -0.14, -0.12, -0.11, -0.09, -0.08, -0.02, 0.00, 0.05, 0.06, 0.06, 0.06, 0.10, 0.10, 0.13, 0.14, 0.26, 0.29, 0.34, 0.36, 0.36, 0.37, 0.38, 0.41, 0.45, 0.45, 0.52, 0.55, 0.56, 0.62, 0.64, 0.65, 0.65, 0.68, 0.71, 0.75, 0.80, 0.81, 0.82, 0.85, 0.86, 0.95, 1.08, 1.08, 1.10, 1.16, 1.22, 1.37, 1.43, 1.44, 2.38, 3.60, 4.10, 5.20, 5.70

	Largest value removed					
Smallest value removed						*5.70*
					5.70	*5.20*
				5.70	*5.20*	*4.10*
			5.70	*5.20*	*4.10*	*3.60*
	None	*5.70*	*5.20*	*4.10*	*3.60*	*2.38*
None	-0.50	-4.15	-7.85	-9.15	-9.70	-2.74
-4.00	1.06	-3.20	-7.65	-9.58	-10.80	-4.27
-4.00, -2.83	4.56	0.02	-4.80	-7.06	-8.62	-2.33
-4.00, -2.83, -2.79	7.99	3.14	-2.09	-4.71	-6.67	-0.65
-4.00, -2.83, -2.79, -2.42	11.88	6.80	1.26	-1.65	-3.92	1.87

3.8 Situation when a large normally distributed data set is used

A normally distributed large data set composed on 500 random observations with a mean of 0 and a standard deviation of 1 was generated. No outliers were added. The proposed procedure indicated that no outlier was present. Table 12 shows the results.

Table 12. Generated data and U_t obtained.

Data (sample size = 500):
Seven smallest values
-3.53, -2.65, -2.53, -2.40, -2.39, -2.32, -2.26, …
Seven largest values
…, 2.44, 2.53, 2.55, 2.56, 2.63, 2.73, 2.79

	Largest value removed				
Smallest value removed					*2.79*
				2.79	*2.73*
			2.79	*2.73*	*2.63*
	None	*2.79*	*2.73*	*2.63*	*2.56*
None	-0.50	3.75	8.10	12.64	17.31
-3.53	1.78	5.97	10.27	14.77	19.39
-3.53, -2.65	6.46	10.63	14.90	19.37	23.97
-3.53, -2.65, -2.53	11.38	15.53	19.77	24.22	28.80
-3.53, -2.65, -2.53, -2.40	16.56	20.69	24.91	29.34	33.90

4. Conclusions

A simple method for the detection of outliers is proposed in this paper and its performance is compared to other techniques already proposed. Future work should test the proposed method employing real data. However, it is expected that the method will perform efficaciously.
The advantages of the simple method are that i) it is easy to calculate, ii) it does not require determining the number of potential outliers in advance, iii) it does not depend on tables or charts, iv) it effectively signals the absence of outliers, and v) it can be used with large sample sizes. Given these benefits, the present method can be easily implemented in current statistical software and be used as an "automatic outlier detection procedure" (see Appendix).
The present procedure can be useful in the study of residual values in regression analysis. This line of research is still to be advanced. Also, future work could systematically study the performance of this procedure via intensive computer simulations (e.g., Monte Carlo studies). [2]

References

[1]. Barnett, V., & Lewis, T. (1994). *Outliers in statistical data (3rd Ed.).* John Wiley & Sons.
[2]. Beckman, R. J., & Cook, R. D. (1983). Outlier.....s. *Technometrics, 25* (2), 119-149.
[3]. Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics, 11* (1), 1-21.
[4]. Johnson, D. E., McGuire, S. A., & Milliken, G. A. (1978). Estimating σ^2 in the presence of outliers. *Technometrics, 20* (4), 441-455.
[5]. Kitagawa, G. (1979). On the use of AIC for the detection of outliers. *Technometrics, 21* (2), 193-199.
[6]. Rosner, B. (1983). Percentage points for a generalised ESD many-outlier procedure. *Technometrics, 25* (2), 165-172.
[7]. Takeuchi, K. (1980). *Statistics numbers in phenomenon and behaviour*. Shincho Sha Publisher. (Original work published in Japanese).
[8]. Takeuchi, K., & Ohashi, Y. (1981). Statistical *inference - 2 Sample problems*. Nihon Hyoronsha Publisher. (Original work published in Japanese).
[9]. Washio, Y., & Ohashi, Y. (1989). *Multidimensional data analysis*. Iwanami Books Publisher. (Original work published in Japanese).

[2] In the original article, under the title of "acknowledgments", the author adds that at first U_t (when n is approximately 10, $n \log \sigma + 2s$) was intuitively derived without any formal evidence and vaguely defined. However, thanks to the author's advisor, the rationale for such formula could be formally defined and tested. Likewise, reviewers' feedback led the author to re-formulate the correction term in relation to sample size. Those suggestions enabled Ueda to furnish an appropriate formula for the correction term.

Appendix[3]

- The Stirling Formula is used to compute log *n*!

```
> stirling = function(x)
{
1/2*log(2*pi)+(1/2+x)*log(x)-x
}
```

- U_t function implemented in *S* programming language and its example:

```
> outlier = function(x, s)
{
# x = data; s = number of outliers
 n = length(x)
 sigma = sqrt(sum((x-mean(x))^2)/n) ###### see footnote [4]
 ut = n*log(sigma)+(sqrt(2)*stirling(n))/n*s
 list(Ut=ut)
}
```

- Example. Takeuchi's (1980) data

```
> takeuchi = c(5.71, 6.57, 7.29, 8.06, 13.32)
> scale(takeuchi)
# the function "scale" scale is generic function whose default method centres and/or
# scales the columns of a numeric matrix. In other words, it gives the z score for each
# observation. Also, "scale" function outputs the mean and standard deviation of the data
# set. The output should look like this
[,1]
[1,] -0.82760285
[2,] -0.54061154
[3,] -0.30033974
[4,] -0.04338241
[5,]  1.71193653
attr(,"scaled:center")
[1] 8.19
attr(,"scaled:scale")
[1] 2.996606 ###### see footnote [5]
> outlier(scale(takeuchi)[1:5],0)
# This line of code outputs the case when no outliers are expected, i.e., n = 5 and s = 0
# the output should look like this
$Ut
[1] -0.5578589
> outlier(scale(takeuchi)[1:4],1)
# This line of code outputs the case when one outlier is expected, i.e., n = 4 and s = 1
# the output should look like this
$Ut
[1] -3.834097
```

[3] In the original article the code was implemented in *S* programming language. Note that this code can be also implemented in *R* programming language. Minor typo changes and re-arrangement of functions were made on the original code so that it can be easily implemented and run in *R* environment. Some comments are added for clarity.

[4] In current versions of *R* the function "sd" could be use to replace "sigma = …". However, the "sd" function uses by default the degrees of freedom adjustment. In terms of the formula used originally by Ueda, the "sd" function would be "sigma = sqrt(sum((x-mean(x))^2)/(n-1))". This could be an issue to bear in mind for future work using Ueda's method. In addition, further investigation could aim to check the performance of Ueda's method when some terms are replaced by robust estimators of location and scale, e.g. median and median average deviation.

[5] The value for the standard deviation given by the function "scale" includes the degrees of freedom correction (see footnote 4). If the formula for sigma proposed by Ueda to calculate the standard deviation (i.e., "sigma = sqrt(sum((x-mean(x))^2)/n)") were used to compute the sigma of Takeuchi's data set, it would be 2.680246.